
Evolutionary conservation in the untranslated regions of actin mRNAs: DNA sequence of a human beta-actin cDNA

Phyllis Ponte, Sun-Yu Ng[§], Joanne Engel⁺, Peter Gunning and Larry Kedes*

MEDIGEN Project, Department of Medicine, Stanford University School of Medicine and Veterans Administration Medical Center, Palo Alto, CA 94304, USA

Received 11 October 1983; Revised and Accepted 13 December 1983

ABSTRACT

We report the complete nucleotide sequence of a human β actin cDNA. Both the 5' and 3' untranslated regions of the sequence are similar (>80%) to the analogous regions of the rat β -actin gene reported by Nudel et al (1983)¹. When a segment of the 3' untranslated region is used as a radiolabelled probe, strong hybridization to chick β actin mRNA is seen. This conservation of sequences suggests that strong selective pressures operate on non-translated segments of β actin mRNA.

INTRODUCTION

We have reported the isolation of cDNA clones encoding the two sarcomeric actins, α skeletal and α cardiac, and the two cytoskeletal actins β and γ ^{2,3}. Partial DNA sequencing of the region encoding diagnostic amino-terminal amino acids established the isotype of each of these clones. Subsequently we constructed subclones derived from the 3'-untranslated (3'-UT) regions of these cDNAs, and demonstrated — by Southern blot analysis and sequence comparisons — that these untranslated regions are isotype specific⁴. However, genomic DNA blotting experiments showed that the 3'-UT regions of all four of these actin mRNAs are conserved in mammals⁴.

Here we present the complete sequence of the human β actin cDNA clone. We extend our observation that the non-translated regions of the actins are well conserved in evolution by comparing the sequences of human β actin with rat β actin¹. Also, by RNA blot analysis, we demonstrate that there is considerable homology between human and chicken β actin 3'-UT regions, in contrast to a previously published result⁵.

METHODS**DNA Sequencing**

DNA sequencing was done by the method of Maxam and Gilbert⁶. Greater than 95% of the sequence shown was done at least twice and/or on opposite strands. Sequence data was managed with the GEL program (IntelliGenetics Inc.)

RNA Preparations and Blot Hybridizations

Brain, heart, liver, and skeletal muscle tissues were dissected from 2 day old chicks. Total RNA was isolated from these samples by the guanidine hydrochloride method followed by phenol/chloroform

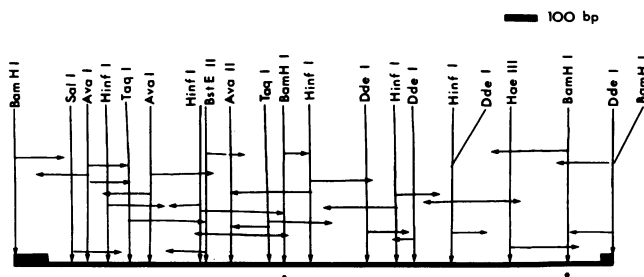


Figure 1 Schematic diagram of the strategy used to determine the DNA sequence of the β actin cDNA insert. Heavy lines indicate vector sequences which flank the cDNA insert, the lighter horizontal line indicates the cDNA insert, which is presented in 5' to 3' orientation. The map shown only indicates the sites used for sequencing and does not represent a complete restriction map of every enzyme indicated. The two *Bam*H I sites indicated by asterisks are present in the vector sequences flanking pHF7 (see Figure 1 of Gunning et.al.²)

extraction as previously described⁷. RNA (10 μ g) was electrophoresed on formaldehyde gels, blot-transferred to nitrocellulose filters and hybridized as described by Thomas⁸. Filters were hybridized to nick-translated⁹ human β actin 3'-UT region DNA (pHF β A-3'UT⁴). Hybridization conditions were 50% formamide, 4 X S.S.C., 50 mM sodium phosphate buffer, pH 6.8, 1.25 X Denhardt's, and 10% dextran sulphate at 42°C. Filters were washed at 0.5 X S.S.C. at 50°C.

RESULTS

The strategy used to determine the sequence of the human β actin cDNA is presented in Figure 1. In addition to sequences derived from pHF β A-1, data was compiled from three shorter clones pHF5 and pHF7², and a third clone whose sequence starts at amino acid 63. There were no sequence differences between the four clones.

The complete sequence of the pHF β A-1 insert is presented in Figure 2. It is 1761 base pairs long and encodes the entire β actin coding region, 41 base pairs of 5'-UT region, and the complete 595 base pair 3'-UT region (which is flanked by a poly(dA) tail of about 50 base pairs). The polyadenylation signal AATAAA¹⁰ occurs at position 1742, 19 base pairs upstream from the poly(dA) tail. The DNA sequence predicts an actin protein which is identical to the human β actin protein sequence determined by Vandekerckhove and co-workers¹¹. (The sequence of the human β actin protein is identical to that of bovine β actin¹¹.)

The base composition of the coding region is characterized by a high G + C content (60.6%). This elevated G + C content in the coding region is attributable to a strong bias in codon usage (Table I) as 84.5% of the third base positions are G or C. Bias towards the use of G or C in this position is also characteristic of other mammalian actin genes: 89% for human α actin, 77.5% for rat α actin, 66.4% for human cardiac actin (all calculations by Hanauer et al.¹²), and 73.6% for rat β actin¹ (our calculations). The high G + C content of the coding region contrasts that of the 3'-UT region which is A + T rich (57.7%).

Inter- and intraspecies comparison of actin genes should furnish insight into the evolution of this complex multi-gene family, and help identify regions responsible for their differential regulation. We have compared the human β actin sequence presented here to the sequence of a rat β actin gene¹, to the sequence of two processed human β actin pseudogenes¹³, and to the partial sequence of a human β actin cDNA derived from a keratinocyte library¹⁴.

Comparison of Human β Actin to Rat β Actin

In Figure 2 we present the comparison of coding region sequences between human and rat β actin genes. Although the two proteins differ at amino acid 15, this difference represents a substitution in the rat protein which is inconsistent with all previous data on actin protein sequences. Therefore, we suspect that this change results from an error in the published rat sequence. Consequently, we have not included these mismatched bases in the calculations which follow.

The human and rat sequences have an overall homology of 79% in the coding region. Most of these base changes (85) are the result of third base substitutions; there are also three first base changes and one second base change which do not affect the amino acid utilized. When calculated by the method of Perler et. al.¹⁵, the per cent of silent site substitutions is 44.4%, after correcting for multiple events. This is less than the 61% silent changes between the human and rat α actin coding sequences¹². This does not necessarily mean that the rates of divergence between α genes and between β genes are different since silent substitutions may not provide a reliable evolutionary clock¹⁵.

Humans and rats diverged approximately 85 million years ago¹⁶. Therefore, the extensive similarity between the DNA sequences of the untranslated regions of the human and rat β actin mRNAs, as seen in Figure 3, is quite intriguing. In the 5'-UT region (Figure 3A), the $K_N(1)$ value (as defined by Miyata et.al.¹⁷) is 0.146 ($K_N(2) = 0.186$). This figure may not reflect the divergence rate for the entire 5'-UT since, for the following reasons, our human β actin cDNA insert appears to be slightly shorter than full length: first, rat β actin mRNA has a 5'-UT region of approximately 80 base pairs¹, and secondly, two reverse transcript type human β actin pseudogenes are highly homologous to each other up to a point 100 base pairs upstream from the initiator methionine¹³.

We are able to compare the complete human and rat sequences for the 3'-UT region since the human cDNA was constructed in a vector which primes first strand synthesis from a covalently linked poly(dT) strand¹⁸, and since the rat sequence was obtained from a genomic clone. The sequence comparison for the 3'-UT region is presented in Figure 3B. It is readily apparent from the figure that the two sequences are highly homologous. The nucleotide differences of the two sequences are fairly evenly distributed over the entire 3'-UT region. This is in contrast to the pattern of conservation observed between the chick and rat α skeletal actin 3'-UT region¹⁹ and the human α actin 3'-UT (our unpublished data). In this three way comparison, the sequence conservation is confined to the 3' half of the 3'-UT region; essentially no homology can be detected in the 5' half of the 3'-UT region.

When the human and rat β actin 3'-UT region sequences are aligned for maximal homology (Figure 3B), $K_N(1)$ is 0.11 (65 of 581 bases mismatched), and $K_N(2) = 0.23$ (154/670). The large difference between

```

      10      20      30      40
TTGCCGATCC GCCGCCGTC CACACCCGCC GCCAGCTCAC C
                                     95
ATG GAT GAT GAT ATC GCC GCG CTC GTC GTC GAC AAC GGC TCC GGC ATG TGC AAG
... ..C ... ..T ... ..G ... ..C ... ..
MET Asp Asp Asp Ile Ala Ala Leu Val Val Asp Asn Gly Ser Gly Met Cys Lys
                                     10
GCC GGC TTC GCG GGC GAC GAT GCC CCC CGG GCC GTC TTC CCC TCC ATC GTG GGG
... ..T ... ..C
Ala Gly Phe Ala Gly Asp Ala Pro Arg Ala Val Phe Pro Ser Ile Val Gly
      20      30
CGC CCC AGG CAC CAG GGC GTG ATG GTG GGC ATG GGT CAG AAG GAT TCC TAT GTG
... ..T ... ..T ... ..T ... ..C ... ..C ...
Arg Pro Arg His Gln Gly Val Met Val Gly Met Gly Gln Lys Asp Ser Tyr Val
      40      50
GGC GAC GAG GCC CAG AGC AAG AGA GGC ATC CTC ACC CTG AAG TAC CCC ATC GAG
... ..G ... ..
Gly Asp Glu Ala Gln Ser Lys Arg Gly Ile Leu Thr Leu Lys Tyr Pro Ile Glu
      60      70
CAC GGC ATC GTC ACC AAC TGG GAC GAC ATG GAG AAA ATC TGG CAC CAC ACC TTC
... ..C ..A ... ..T ... ..G ..T ... ..T ...
His Gly Ile Val Thr Asn Trp Asp Asp Met Glu Lys Ile Trp His His Thr Phe
      80      90
TAC AAT GAG CTG CGT GTG GCT CCC GAG GAG CAC CCC GTG CTG CTG ACC GAG GCC
... ..C ..T ... ..T ... ..C ... ..
Tyr Asn Glu Leu Arg Val Ala Pro Glu Glu His Pro Val Leu Leu Thr Glu Ala
      100
CCC CTG AAC CCC AAG GCC AAC CGC GAG AAG ATG ACC CAG ATC ATG TTT GAG ACC
... ..T ... ..T ... ..C ..T ..A ... ..
Pro Leu Asn Pro Lys Ala Asn Arg Glu Lys Met Thr Gln Ile Met Phe Glu Thr
      110      120
TTC AAC ACC CCA GCC ATG TAC GTT GCT ATC CAG GCT GTG CTA TCC CTG TAC GCC
... ..C ... ..A ... ..G ... ..
Phe Asn Thr Pro Ala Met Tyr Val Ala Ile Gln Ala Val Leu Ser Leu Tyr Ala
      130      140
TCT GGC CGT ACC ACT GGC ATC GTG ATG GAC TCC GGT GAC GGG GTG ACC CAC ACT
... ..T ... ..T ... ..T ... ..A ... ..
Ser Gly Arg Thr Thr Gly Ile Val Met Asp Ser Gly Asp Gly Val Thr His Thr
      150      160
GTG CCC ATC TAC GAG GGG TAT GCC CTC CCC CAT GCC ATC CTG CGT CTG GAC CTG
... ..T ... ..T ..C ..G ... ..T ... ..
Val Pro Ile Tyr Glu Gly Tyr Ala Leu Pro His Ala Ile Leu Arg Leu Asp Leu
      170
GCT GGC CGG GAC CTG ACT GAC TAC CTC ATG AAG ATC CTC ACC GAG CGC GGC TAC
... ..A ... ..G ... ..T ... ..
Ala Gly Arg Asp Leu Thr Asp Tyr Leu Met Lys Ile Leu Thr Glu Arg Gly Tyr
      180      190
AGC TTC ACC ACC ACG GCC GAG CGG GAA ATC GTG CGT GAC ATT AAG GAG AAG CTG
... ..A ..T ... ..A ... ..A ... ..
Ser Phe Thr Thr Thr Ala Glu Arg Glu Ile Val Arg Asp Ile Lys Glu Lys Leu
      200
TGC TAC GTC GCC CTG GAC TTC GAG CAA GAG ATG GCC ACG GCT GCT TCC AGC TCC
... ..T ..T ... ..A ... ..T ... ..C ..A ... ..T ...
Cys Tyr Val Ala Leu Asp Phe Glu Gln Glu Met Ala Thr Ala Ala Ser Ser Ser
      210      220      230

```

```

797
TCC CTG GAG AAG AGC TAC GAG CTG CCT GAC GGC CAG GTC ATC ACC ATT GGC AAT
... ..T ... ..T ... ..C ... ..
Ser Leu Glu Lys Ser Tyr Glu Leu Pro Asp Gly Gln Val Ile Thr Ile Gly Asn
234a 240 250 851
GAG CGG TTC CGC TGC CCT GAG GCA CTC TTC CAG CCT TCC TTC CTG GGC ATG GAG
... ..A ... ..C ... ..T ... ..T ... ..T ... ..A
Glu Arg Phe Arg Cys Pro Glu Ala Leu Phe Gln Pro Ser Phe Leu Gly Met Glu
260 905
TCC TGT GGC ATC CAC GAA ACT ACC TTC AAC TCC ATC ATG AAG TGT GAC GTG GAC
... ..T ... ..A ... ..T ... ..T ... ..T ... ..T
Ser Cys Gly Ile His Glu Thr Thr Phe Asn Ser Ile Met Lys Cys Asp Val Asp
270 280 959
ATC CGC AAA GAC CTG TAC GCC AAC ACA GTG CTG TCT GGC GGC ACC ACC ATG TAC
... ..T ... ..C ... ..T ... ..T ... ..T ... ..T
Ile Arg Lys Asp Leu Tyr Ala Asn Thr Val Leu Ser Gly Gly Thr Thr Met Tyr
290 300 969
CCT GGC ATT GCC GAC AGG ATG CAG AAG GAG ATC ACT GCC CTG GCA CCC AGC ACA
..A ... ..C ..T ... ..T ... ..T ... ..T ... ..T ..C
Pro Gly Ile Ala Asp Arg Met Gln Lys Glu Ile Thr Ala Leu Ala Pro Ser Thr
310 320 1013
ATG AAG ATC AAG ATC ATT GCT CCT CCT GAG CGC AAG TAC TCC GTG TGG ATC GGC
Met Lys Ile Lys Ile Ile Ala Pro Pro Glu Arg Lys Tyr Ser Val Trp Ile Gly
330 340 1067
GGC TCC ATC CTG GCC TCG CTG TCC ACC TTC CAG CAG ATG TGG ATC AGC AAG CAG
... ..T ... ..A ... ..T ... ..T ... ..T ... ..T ... ..T
Gly Ser Ile Leu Ala Ser Leu Ser Thr Phe Gln Gln Met Trp Ile Ser Lys Gln
350 1121
GAG TAT GAC GAG TCC GGC CCC TCC ATC GTC CAC CGC AAA TGC TTC TAG G
... ..C ..T ... ..T ... ..T ... ..T ... ..T ... ..T
Glu Tyr Asp Glu Ser Gly Pro Ser Ile Val His Arg Lys Cys Phe
370 374 1166 1170
CGGACTATGA CTTAGTTGCG TTACACCCTT TCTTGACAAA ACCTAAGTTG CGCAGAAAAC
1230
AAGATGAGAT TGGCATGGCT TTATTTGTTT TTTTGTGTTT GTTTTGTTT TTTTGTGTTT
1290
TTTGGCTTGA CTCAGGATTT AAAAAGTGA ACGGTGAAGG TGACAGCAGT CGGTTGGAGC
1350
GAGCATCCCC CAAAGTTCAC AATGTGGCCG AGGACTTTGA TTGCACATTG TTGTTTTTTT
1410
AATAGTCATT CCAAATATGA GATGCATTGT TACAGGAAGT CCCTTGCCAT CCTAAAAGCC
1470
ACCCCACTTC TCTCTAAGGA GAATGGCCCA GTCCTCTCCC AAGTCCACAC AGGGGAGGTG
1530
ATAGCATTCG TTTCGTGTAA ATTATGTAAT GCAAAATTTT TTTAATCTTC GCCTTAATAC
1590
TTTTTTATTT TTTTTTATTT TGAATGATGA GCCTTCGTGC CCCCCCTTCC CCCITTTTGT
1650
CCCCCAACTT GAGATGTATG AAGGCTTTTG GTCTCCCTGG GAGTGGGTGG AGGCAGCCAG
1710
GGCTTACCTG TACACTGACT TGAGACCAGT TGAATAAAG TGACACACCTT A
1761

```

Figure 2 Sequence of human β actin cDNA. The upper line is the sequence of the human β actin cDNA, numbered consecutively from the 5' end. The line immediately beneath presents the comparison with the rat β actin, in the coding region only. Only base changes between the two sequences are marked. The bottom line shows the amino acid sequence, numbered according to the convention of Lu and Elizinga²¹, which is predicted from the human cDNA.

TABLE I. CODON USAGE IN HUMAN β ACTIN

TTT-Phe 1 (.3)	TCT-Ser 2 (.5)	TAT-Tyr 3 (.8)	TGT-Cys 2 (.5)
TTC-Phe 12 (3.2)	TCC-Ser 16 (4.3)	TAC-Tyr 12 (3.2)	TGC-Cys 4 (1.1)
TTA-Leu 0 (0)	TCA-Ser 0 (0)	TAA- . 0 (0)	TGA- . 0 (0)
TTG-Leu 0 (0)	TCG-Ser 1 (.3)	TAG- . 0 (0)	TGG-Trp 4 (1.1)
CTT-Leu 0 (0)	CCT-Pro 6 (1.6)	CAT-His 1 (.3)	CGT-Arg 4 (1.1)
CTC-Leu 6 (1.6)	CCC-Pro 12 (3.2)	CAC-His 8 (2.1)	CGC-Arg 7 (1.9)
CTA-Leu 1 (.3)	CCA-Pro 1 (.3)	CAA-Gln 1 (.3)	CGA-Arg 0 (0)
CTG-Leu 20 (5.3)	CCG-Pro 0 (0)	CAG-Gln 11 (2.9)	CGG-Arg 4 (1.1)
ATT-Ile 4 (1.1)	ACT-Thr 5 (1.3)	AAT-Asn 2 (.5)	AGT-Ser 0 (0)
ATC-Ile 24 (6.4)	ACC-Thr 17 (4.5)	AAC-Asn 7 (1.9)	AGC-Ser 6 (1.6)
ATA-Ile 0 (0)	ACA-Thr 2 (.5)	AAA-Lys 3 (.8)	AGA-Arg 1 (.3)
ATG-MET 17 (4.5)	ACG-Thr 2 (.5)	AAG-Lys 16 (4.3)	AGG-Arg 2 (.5)
GTT-Val 1 (.3)	GCT-Ala 7 (1.9)	GAT-Asp 5 (1.3)	GGT-Gly 2 (.5)
GTC-Val 8 (2.1)	GCC-Ala 18 (4.8)	GAC-Asp 18 (4.8)	GGC-Gly 23 (6.1)
GTA-Val 0 (0)	GCA-Ala 2 (.5)	GAA-Glu 2 (.5)	GGA-Gly 0 (0)
GTG-Val 13 (3.5)	GCG-Ala 2 (.5)	GAG-Glu 24 (6.4)	GGG-Gly 3 (.8)

$K_N(1)$ and $K_N(2)$ derives from the fact that the rat 3'-UT is longer than the human 3'-UT (664 base pairs based on S1 mapping experiments vs 595 base pairs). Most of the length difference can be accounted for by three gaps of 34, 11, and 7 base pairs in the human sequence. Therefore, the $K_N(1)$ may be a more accurate predictor of inter-species hybridization behavior of probes derived from this region, since stable hybrids of lengths considerably less than the full length of the region should occur.

If no selective pressures are operating on the 3'-UT region then we would expect a K_N of approximately 0.6, assuming a rate of divergence of 0.7% basechanges per myr for random drift¹⁵. Thus the striking similarity of the rat and human β actin genes in the 3'-UT regions suggests that there is considerable evolutionary pressure to maintain the sequence.

Conservation of Actin Untranslated Regions

The conservation of sequence between the 3'-UT regions of human and rat β actin mRNAs suggested to us that the 3'-UT regions of β actin mRNAs of even more distantly related species might share considerable sequence similarity. Accordingly we assayed for the cross-hybridization of the human β actin 3'-UT region probe to chicken RNA. Total cellular RNA was isolated from the brain, liver, heart, and skeletal muscle of 2 day old chicks. 10 μ g of each RNA sample was electrophoresed on a formaldehyde gel, and the RNA was transferred to nitrocellulose by blotting. The filter was hybridized to human β actin 3'-UT region probe (pHF β A-3'UT) under the conditions described in Methods. The resulting autoradiogram is shown in Figure 4. A strong hybridization signal occurs at the mobility of cytoplasmic actin mRNA in lanes containing chick brain (B) and liver (L) RNA. No hybridization is seen in lanes containing heart (H) or skeletal muscle (M) RNA, even though muscle actin mRNAs are among the most abundant mRNAs present in these tissues. This data is consistent with the interpretation that the human β actin 3'-UT region probe is hybridizing exclusively to chicken β actin mRNA.

This result is somewhat unexpected since Cleveland and co-workers have argued that the 3'-UT regions of actin genes diverge rapidly⁵. Their conclusion was based on their inability to detect hybridization of

A. 5'-UT REGION HOMOLOGY : HUMAN VS RAT

```

TTGCCGATCCGCCG-CCCG-TCCACACCCGCCGCCAGCTCACC
  *   *   *   *   *   *   *   *   *   *   *
TTGCAGCTCCTCCGTCCCGTCCACACCCGCCACCAGTTCGCC

```

B. 3'-UT REGION HOMOLOGY : HUMAN VS RAT

```

TAGGCGGACTATGACTTAGTTGCGTT--ACACCCCTTCTT-GACAAACCTAACTTGGCGAGAAAAAAG      1233
  *   *   *   *   *   *   *   *   *   *   *
TAGGCGGACTGTTACTGAGCTGCGTTTTACACCCCTTCTTGGACAAACCTAACTTGGCGAGAAAAAAG

-ATGAGA---TTGGCATGGCTTTATTTGTTTTTTTTGTTTT--GTTTTGGTTTTTTTTT-----      1286
  *   *   *   *   *   *   *   *   *   *   *
AATGAGACATTTGGCATGGCTTTATT-GTTTTTTT-GTTTTTTGTTTTGTTTTTTTTTAAATTTTTTTTT

-----TTTTTTTTTGGC-----TTGACTCA-GGATTIAAAACCTGGAACGGTGAA      1328
  *   *   *   *   *   *   *   *   *   *   *
TAAAAAGGTTTTTTTTTTTTGTTTGTGTTTGGCGCTTTTGACTCAAGGATTAACAACTGGAACGGTGAA

GGTGACAGCAGTCGGTTGGAGCGAGCATCCCCAAAGTTC-ACAATGTGGCCGAGGACTTTGATTGCACA      1397
  *   *   *   *   *   *   *   *   *   *   *
GGCGACCGCAGTTGGTTGGAGCAACATCCCCAAAGTTCACAATGTGGCTGAGGACTTTGATTGTACA

TTGTT-----GTTTTTTTAATAGTCATTCCAATAT----GAGATGCATTGTTACAGGAAGTCCC      1453
  *   *   *   *   *   *   *   *   *   *   *
TTGTTTTTTGTTTTTGGTTTTTTAATAGTCACTCAAGTATCCACGGCATAGATGGTACAGGAAGTCCC

TTGCCATCTAAAGCCACCCCACTTCTCTCTAAGG--AGAATGGCCAGTCC-TCTCCCAAGTCCACACA      1521
  *   *   *   *   *   *   *   *   *   *   *
TCACCTCCCAAAGCCACCCC-CAACTC-CTAAGGGGAGGATGGCTGCATCCATGCCCTGAGTCCACACC

GGGGAGGTGATAGCAATTGCTTTCGTGTAAATTATGTA-ATGCAAA-ATTTTTTTAA-TCCTC-CGCCTTAA      1587
  *   *   *   *   *   *   *   *   *   *   *
GGGAAGGTGACAGCATGCTTCTGTGTAAATTATGTAATGCAACATTTTTTTAAATCTCCCGCCTTAA

TACTTTTTTATTTTGTATTTT-TGAATGATGAGCC-TTCGTGCCCCCTTCCCCCTTTTGT-CCCCC      1655
  *   *   *   *   *   *   *   *   *   *   *
TACTTCATTTTGTTTTAAATTTCTGAATGGTCAGCCATTCGTGCCCC-----TTTTTTTTGTGCCCCC

AACTTGAGATGATGAAGGCTTTTGGTCTCCCTGGGAGTGGGTGGAGGCAGCCAGGGCTTACCTGTACACT      1726
  *   *   *   *   *   *   *   *   *   *   *
AACTT--GATGTATGAAGGCTTT-GGTCTCCCTGGGAGGTGT-GAGGC-GCCAGGGCTGGCCTGTACACT

GACTTGAGACCAGTTGAATAAAAGTGACACCTTA      1761
  *   *   *   *   *   *   *   *   *   *   *
GACGTGAGACCGTTTAAATAAAAGTGACACCTTA

```

Figure 3 Comparison of the 5'- and 3'-UT regions of the human β actin cDNA to the rat β actin gene. The sequences have been aligned for maximal homology. The human sequence is numbered as in Figure 2. Upper = human; lower = rat. A. 5'-UT region. B. 3'-UT region. Mismatches and gaps are indicated by asterisks.

subcloned chick β and γ actin 3'-UT region probes to genomic DNA of other species under conditions (50% formamide, 3 X S.S.C., 41°C.) in which actin coding region probes cross hybridize strongly. We assume that the differences in our experiments lies in nature of the probes and hybridization conditions used. Yet while

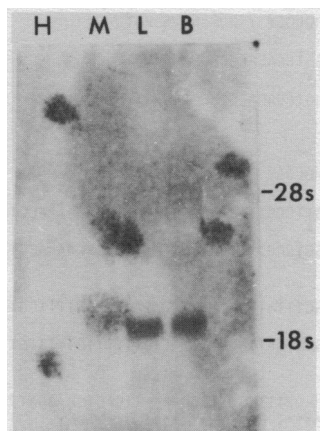


Figure 4 Hybridization of human β actin 3'-UT region probe to chick RNA. 10 μ g of total cellular RNA from chick heart (H), skeletal muscle (M), liver (L), and brain (B) was electrophoresed on a formaldehyde gel. RNA was transferred to nitrocellulose and the filter was hybridized to human β actin 3'-UT region probe (pHF β A-3'UT⁴). The resulting autoradiogram shows hybridization to actin mRNA in liver and brain, but not in heart and skeletal muscle, as expected for β actin. The position of chick ribosomal RNAs are indicated.

we concur that the 3'-UT regions of actin genes are isotype specific, we conclude that actin gene 3'-UT regions show convincing evidence of strong sequence conservation.

Generation of β Actin Pseudogenes

We have previously reported that the sarcomeric actin genes are single copy in the genome whereas the cytoskeletal actin genes are represented by many copies, the majority of which appear to be pseudogenes⁴. This data suggests that the fixation of large numbers of reverse-transcript type ("processed") pseudogenes is a function of germ cell expression. The partial sequence of clones containing two processed human β actin pseudogenes has been published¹³. We have compared the sequence of the human β actin cDNA to these pseudogenes in order to estimate the time at which they were fixed in the genome. Since the complete 3'-UT region of the pseudogenes was not determined, only the actin coding regions could be compared. The coding region of H β Ac- Ψ 1 differs from the human β actin cDNA coding sequence at 120 of 1125 positions (112 base changes, 6 deletions and 2 insertions, $K_N(1) = 0.11$). H β Ac- Ψ 2 differs in 135 of 1125 positions (103 base changes, 26 deletions and 6 insertions; $K_N(1) = 0.092$). Using 0.7% base changes per myr as the rate of fixation of neutral mutations in the genome¹⁵, we estimate that these genes arose approximately 15.7 and 13.1 myr ago, respectively. This estimate assumes that the two pseudogenes accumulate mutations randomly and that no correction mechanisms are operative.

Transcription of Human β Actin Genes

Hanukoglu and Fuchs have characterized an 819 base pair β actin cDNA clone which was isolated from a human keratinocyte library¹⁴. The sequence of this cDNA is identical to the one presented here except for

one nucleotide difference, a third base change at amino acid 266. This difference may result from a sequencing error or may represent a distinct allele in the human population. The presence of only one base difference in 819 bases suggests that the same β actin gene is being utilized in two human cells of different germ layer origins. The possible significance of this observation has been discussed in detail elsewhere⁴.

DISCUSSION

We have presented the complete sequence of a human β actin cDNA clone which was isolated from an SV-40 transformed fibroblast library. The amino acid sequence predicted from the cDNA sequence is identical to the human (and bovine) β actin protein sequence¹¹. In agreement with our previous conclusion based on DNA blot results⁴, a comparison of the human β actin sequence to the rat β actin sequence¹ demonstrates that there is considerable homology between the untranslated regions of the two genes. Although the untranslated regions flanking actin mRNAs are isotype specific, they are not species specific. Thus, at least in the vertebrates, there is a very much slower rate of divergence of the 5'- and 3'-untranslated regions of actin genes than was previously thought to occur.

This conclusion is supported by the ability of a cloned human actin 3'-UT region probe to hybridize to an actin mRNA present in chicken liver and brain, but not to the actin mRNA present in chick heart and skeletal muscle. Although this data conflicts with the DNA blot data of Cleveland et al.⁵, we feel that the two sets of data can be reconciled. We estimate, using standard formulas, that in both cases the hybridization conditions used would be near to the T_m of duplexes with an A+T content of 60% and 20% mismatch. However, since DNA:DNA hybrids are less stable in formamide than DNA:RNA hybrids, duplexes may not have formed under the Southern blot conditions used by Cleveland and coworkers. Shani and coworkers²⁰ reached similar conclusions regarding the rapid divergence of actin untranslated regions when they failed to detect cross-hybridization of their cloned rat α actin 3'-UT region probe to chicken skeletal muscle RNA. However, DNA sequence analysis by Ordahl and Cooper¹⁹ has shown that there is indeed striking similarity between chick and rat α skeletal actin in the 3' half of the 3'-UT region. Similarly, we find that the 3' half of the 3'-UT region of human α skeletal actin is little diverged from either rat or chicken, although probes made from this region cross-hybridize poorly to the other species due to the high A+T content of the region (unpublished observations).

The sequence similarities between the rat, chick and human α actin 3'-UT regions is confined to the 3' half of the untranslated region. This pattern of conservation was previously noted in other genes¹⁷. However, in the comparison presented here, between human and rat β actin, the homology is evenly distributed over the entire 3'-UT region. The extent of conservation, however, is similar to that seen for other human/rodent comparisons. Miyata and coworkers have calculated that for β globin, preproinsulin, and growth hormone genes, the rates of divergence of the 3' portion of the 3'-UT region are 0.196, 0.191, and 0.152, respectively¹⁷. The K_N for the entire human/rat β actin 3'-UT region is 0.11. This evidence of strong evolutionary pressure to maintain the sequence of 3'-UT regions suggests that they may play some valuable role in the expression of actin genes.

ACKNOWLEDGEMENTS

We thank Mike Crowe for the gift of chicks. Adrian Minty contributed valuable comments and critical review of the manuscript. This research was supported by grants to L.K. from the National Institutes of Health, the Veterans Administration and the Muscular Dystrophy Association. P.P. is the recipient of a Research Service Award from the National Institutes of Health. J.E. was a predoctoral fellow of the Medical Scientist Training Program of Stanford Medical School.

*To whom reprint requests should be addressed

§Current Address: Linus Pauling Institute of Biology and Medicine, Page Mill Road, Palo Alto, CA, USA

+Current Address: Department of Medicine, Hospital of the University of Pennsylvania, Philadelphia, PA, USA

References

1. Nudel, U., Zakut, R., Shani, M., Neuman, S., Levy, Z. and Yaffe, D., *Nucleic Acids Res.*, Vol. 11, 1983, pp. 1759-1771.
2. Gunning, P., Ponte, P., Okayama, H., Engel, J., Blau, H. and Kedes, L., *Mol. Cell. Biol.*, Vol. 3, 1983, pp. 787-795.
3. Gunning, P., Ponte, P., Blau, H. and Kedes, L., *Mol. Cell. Biol.*, Vol. 3, 1983, pp. in press.
4. Ponte, P., Gunning, P., Blau, H. and Kedes, L., *Mol. Cell Biol.*, 1983, ; in press
5. Cleveland, D.W., Lopata, M.A., MacDonald, R.J., Cowan, D.J., Rutter, W.J. Kirschner, M.W., *Cell*, Vol. 20, 1980, pp. 95-105.
6. Maxam, A., and Gilbert W., *Proc. Nat. Acad. Sci. USA*, Vol. 74, 1977, pp. 560-564.
7. Childs, G., Maxson, R. and Kedes, L.H., *Dev. Biol.*, Vol. 73, 1979, pp. 153-173.
8. Thomas, P., *Proc. Nat. Acad. Sci. USA*, Vol. 77, 1980, pp. 5201-5205.
9. Rigby, P. W., Dieckmann, M. Rhodes, C., Berg, P., *J. Mol. Biol.*, Vol. 113, 1977, pp. 237-251.
10. Proudfoot, N.J. and Brownlee, G.G., *Nature*, Vol. 263, 1976, pp. 211-214.
11. Vandekerckhove, J., Leavitt, J., Kakunaga, T. and Weber, K., *Cell*, Vol. 22, 1980, pp. 893-899.
12. Hanauer, A., Levin, M., Heilig, R., Daegelon, D., Kahn, A. and Mandel, J.L., *Nucleic Acids Res.*, Vol. 11, 1983, pp. 3503-3516.
13. Moos, M. and Gallwitz, D., *EMBO Journal*, Vol. 2, 1983, pp. 757-761.
14. Hanukoglu, I., Tanese, N. and Fuchs, E., *J. Mol. Biol.*, Vol. 163, 1983, pp. 673-678.
15. Perler, F., Efstratiadis, A., LoMedico, P., Gilbert, W., Kolodner, R. and Dodgson, J., *Cell*, Vol. 20, 1980, pp. 555-566.
16. Romero-Herrera, A.E., Lehmann, H., Joysey, K.A. and Friday, A.E., *Nature*, Vol. 246, 1973, pp. 389-395.
17. Miyata, T., Yasunaga, T. and Nishida, T., *Proc. Natl. Acad. Sci. USA*, Vol. 77, 1980, pp. 7328-7332.
18. Okayama, H. and Berg, P., *Mol. Cell Biol.*, Vol. 3, 1983, pp. 280-289.
19. Ordahl, C.P. and Cooper, T.A., *Nature*, Vol. 303, 1983, pp. 348-349.
20. Shani, M., Nudel, U., Zevin-Sonkin, D., Zakut, R., *Nuc. Acid. Res.*, Vol. 9, 1981, pp. 579-589.
21. Lu, R. C., and Elzinga, M., *Biochem.*, Vol. 16, 197, pp. 5801-5806.